REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	New Reprint	-

4. TITLE AND SUBTITLE

Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation

5a. CONTRACT NUMBER
W911NF-12-1-0552

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

David E. Weinberg, Premal Shah, Stephen W. Eichhorn, Jeffrey A. Hussmann, Joshua B. Plotkin, David P. Bartel

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

University of Pennsylvania
Office of Research Services
3451 Walnut Street, Suite P-221
Philadelphia, PA 19104 -6205

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
62345-MA.16

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Ribosome-footprint profiling provides genome-wide snapshots of translation, but technical challenges can confound its analysis. Here, we use improved methods to obtain ribosome-footprint profiles and mRNA abundances that more faithfully reflect gene expression in Saccharomyces cerevisiae. Our results support proposals that both the beginning of coding regions and codons matching rare tRNAs are more slowly translated. They also indicate that emergent polypeptides with as few as three basic residues within a ten-residue window tend to slow translation. With the improved mRNA measurements, the variation attributable to translational control in

15. SUBJECT TERMS

ribosome profiling

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Joshua Plotkin
UU	UU	UU	UU		19b. TELEPHONE NUMBER 215-898-7133

Report Title

Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation

ABSTRACT

Ribosome-footprint profiling provides genome-wide snapshots of translation, but technical challenges can confound its analysis. Here, we use improved methods to obtain ribosome-footprint profiles and mRNA abundances that more faithfully reflect gene expression in Saccharomyces cerevisiae. Our results support proposals that both the beginning of coding regions and codons matching rare tRNAs are more slowly translated. They also indicate that emergent polypeptides with as few as three basic residues within a ten-residue window tend to slow translation. With the improved mRNA measurements, the variation attributable to translational control in exponentially growing yeast was less than previously reported, and most of this variation could be predicted with a simple model that considered mRNA abundance, upstream open reading frames, cap-proximal structure and nucleotide composition, and lengths of the coding and 5? UTRs. Collectively, our results provide a framework for executing and interpreting ribosome-profiling studies and reveal key features of translational control in yeast.

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)

Continuation for Block 13

ARO Report Number 62345.16-MA
Improved Ribosome-Footprint and mRNA Meas...

Block 13: Supplementary Note

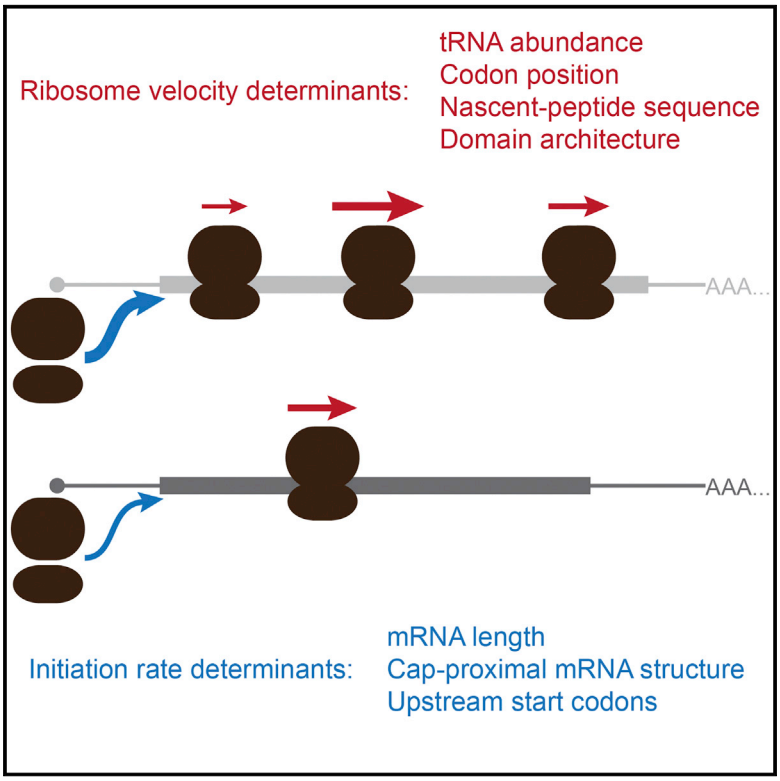
Approved for public release; distribution is unlimited.

Cell Reports

Resource

Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation

Graphical Abstract



Authors

David E. Weinberg, Premal Shah, Stephen W. Eichhorn, Jeffrey A. Hussmann, Joshua B. Plotkin, David P. Bartel

Correspondence

david.weinberg@ucsf.edu

In Brief

Ribosome-footprint profiling provides genome-wide snapshots of translation, but technical challenges can confound its analysis. Here, Weinberg et al. use improved methods to obtain ribosome-footprint profiles and mRNA abundances that together reveal key features of translational control in yeast.

Highlights

- Ribosome dwell times are typically longer at codons for less abundant tRNAs
- The first 200 codons tend to be translated more slowly, regardless of codon choice
- Accurate yeast mRNA abundances reveal a narrow range of translational efficiencies
- A statistical model explains much of the variation in initiation efficiencies

Accession Numbers

GSE75897

Weinberg et al., 2016, Cell Reports *14*, 1787–1799
February 23, 2016
http://dx.doi.org/10.1016/j.celrep.2016.01.043

CellPress

Cell Reports
Resource

Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation

David E. Weinberg,[1,8,*] Premal Shah,[2,8,9] Stephen W. Eichhorn,[3,4,5] Jeffrey A. Hussmann,[6,7] Joshua B. Plotkin,[2] and David P. Bartel[3,4,5]

[1]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA
[2]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[3]Howard Hughes Medical Institute
[4]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[5]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Institute for Computational Engineering and Sciences, University of Texas, Austin, TX 78712, USA
[7]Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA
[8]Co-first author
[9]Present address: Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA
*Correspondence: david.weinberg@ucsf.edu
http://dx.doi.org/10.1016/j.celrep.2016.01.043

SUMMARY

Ribosome-footprint profiling provides genome-wide snapshots of translation, but technical challenges can confound its analysis. Here, we use improved methods to obtain ribosome-footprint profiles and mRNA abundances that more faithfully reflect gene expression in *Saccharomyces cerevisiae*. Our results support proposals that both the beginning of coding regions and codons matching rare tRNAs are more slowly translated. They also indicate that emergent polypeptides with as few as three basic residues within a ten-residue window tend to slow translation. With the improved mRNA measurements, the variation attributable to translational control in exponentially growing yeast was less than previously reported, and most of this variation could be predicted with a simple model that considered mRNA abundance, upstream open reading frames, cap-proximal structure and nucleotide composition, and lengths of the coding and 5′ UTRs. Collectively, our results provide a framework for executing and interpreting ribosome-profiling studies and reveal key features of translational control in yeast.

INTRODUCTION

Although most cellular mRNAs use the same translation machinery, the dynamics of translation can vary between mRNAs and within mRNAs, often with functional consequences. For example, strong secondary structure within the 5′ UTR of an mRNA can impede the scanning ribosome, thereby reducing the rate of protein synthesis (Kozak, 1986; Andersson and Kurland, 1990; Bulmer, 1991; Kudla et al., 2009; Tuller et al., 2010, 2011; Plotkin and Kudla, 2011; Ding et al., 2012; Bentele et al., 2013). The accessibility of the 5′ cap (Godefroy-Colburn et al., 1985; Richter and Sonenberg, 2005) and the presence of small open reading frames (ORFs) within 5′ UTRs referred to as upstream ORFs (uORFs) (Kozak, 1986; Ingolia et al., 2009; Brar et al., 2012; Zur and Tuller, 2013) can also modulate the rate of translation initiation (Sonenberg and Hinnebusch, 2009). Likewise, codon choice, mRNA structure, and the identity of the nascent polypeptide can influence elongation rates (Varenne et al., 1984; Brandman et al., 2012). In addition, differences in elongation rates can influence co-translational protein folding, localization of the mRNA or protein, and in extreme cases the rate of protein production (Kimchi-Sarfaty et al., 2007; Xu et al., 2013; Zhou et al., 2013). Finally, stop-codon readthrough can introduce alternative C-terminal regions that affect protein stability, localization, or activity (Dunn et al., 2013). Despite known examples of regulation at each of these stages of translation, translation is largely controlled at initiation, which is rate limiting for most mRNAs (Andersson and Kurland, 1990; Bulmer, 1991; Chu and von der Haar, 2012; Shah et al., 2013).

Variation in protein abundances observed in yeast cells largely reflects variation in mRNA abundances, indicating that much of gene regulation occurs at the level of mRNA synthesis and decay (Greenbaum et al., 2003; Csárdi et al., 2015). However, differences in translation rates also contribute. Studies using microarrays for global polysome profiling indicate that ribosome densities for different mRNAs vary over a 100-fold range (from 0.03 to 3.3 ribosomes per 100 nucleotides), indicating extensive translation control in *Saccharomyces cerevisiae* (Arava et al., 2003). More recently, the use of ribosome-footprint profiling has enabled transcriptome-wide analyses of translation using high-throughput sequencing, which again suggested a nearly 100-fold range of translational efficiencies (TEs) in log-phase yeast (Ingolia et al., 2009).

The ribosome-profiling method has itself undergone refinements over the last few years. Here, we build upon these advances and present improved ribosome-profiling and mRNA sequencing (mRNA-seq) datasets for log-phase yeast. Comparisons to many previous datasets reveal protocol-specific biases that can influence interpretation of ribosome-profiling experiments. With these insights, we then address several classical questions and ongoing debates in protein translation, such as the influence of tRNA abundances and nascent-peptide sequence on elongation rates. Our improved datasets also constrict the differences in TEs observed in log-phase yeast, such that the gene-to-gene variability that does remain can be largely predicted using a simple statistical model that considers only six features of the mRNAs.

RESULTS

Less Perturbed Ribosome Footprints

Protocols for analyzing polysome profiles or capturing ribosome footprints (referred to as ribosome-protected fragments, or RPFs) typically involve treating cells with the elongation inhibitor cycloheximide (CHX) to arrest the ribosomes prior to harvest (Ingolia et al., 2009; Gerashchenko et al., 2012; Zinshteyn and Gilbert, 2013; Artieri and Fraser, 2014; McManus et al., 2014). An advantage of CHX pre-treatment is that it prevents the run-off of ribosomes that can otherwise occur during harvesting (Ingolia et al., 2009). However, this treatment can also have some undesirable effects. Because CHX does not inhibit translation initiation or termination, pre-treatment of cultures leads to ribosome accumulation at start codons and depletion at stop codons (Ingolia et al., 2011; Guydosh and Green, 2014; Pelechano et al., 2015). In addition, because CHX binding to the 80*S* ribosome is both non-instantaneous and reversible, the kinetics of CHX binding and dissociation presumably allow newly initiated ribosomes to translocate beyond the start codon. Another possible effect of CHX treatment is that ribosomes might preferentially arrest at specific codons that do not necessarily correspond to codons that are more abundantly occupied by ribosomes in untreated cells. Although effects of CHX pre-treatment have minimal consequence for analyses performed at the gene level, i.e., comparisons of the same gene in different conditions, or comparisons between different genes after discarding reads in the 5′ regions of ORFs, CHX pre-treatment may have severe consequences for analyses that require single-codon resolution.

The potential effects of CHX pre-treatment near the start codon have been discussed since the introduction of ribosome profiling, where an alternative protocol with flash-freezing and no CHX pre-treatment is also presented (Ingolia et al., 2009). Indeed, many recent ribosome-profiling experiments avoid CHX pre-treatment (Gardin et al., 2014; Gerashchenko and Gladyshev, 2014; Guydosh and Green, 2014; Jan et al., 2014; Lareau et al., 2014; Pop et al., 2014; Williams et al., 2014; Nedialkova and Leidel, 2015). However, consensus on the ideal protocol has not yet been reached, in part because the influence of alternative protocols on the interpretation of translation dynamics has not been systematically analyzed.

Here, we implemented a filtration and flash-freezing protocol to rapidly harvest yeast cultures. Importantly, this protocol minimized the time the cells experience starvation, which leads to rapid ribosome run-off (Ingolia et al., 2009; Gardin et al., 2014; Guydosh and Green, 2014). The protocol did include CHX in the lysis buffer to inhibit elongation that might occur during RNase digestion, although we doubt this precaution was necessary.

The original ribosome-profiling protocol also used cDNA circularization (Ingolia et al., 2009), while some subsequent protocols instead ligate to a second RNA adaptor prior to cDNA synthesis (Guo et al., 2010). Both approaches can introduce sequence-specific biases at the 5′ ends of reads, which are not expected to influence results of analyses performed at the level of whole mRNAs but might influence results of codon-resolution analyses. Borrowing from methods developed for small-RNA sequencing (Jayaprakash et al., 2011; Sorefan et al., 2012), we minimized these biases by ligating a library of adaptor molecules that included all possible sequences at the eight nucleotides nearest to the ligation junction. Using this ligation protocol with a rapidly harvested, flash-frozen sample, we generated 74.3 million RPFs for log-phase yeast.

The 5′ Ramp of Ribosomes

Using the 5′ ends of RPFs, we inferred the codon at the A site of each footprint (Ingolia et al., 2009). Analysis of all mapped reads revealed the expected three-nucleotide periodicity along the ORFs, as well as ribosome accumulation at the start and stop codons (Figures 1A and 1B).

To examine the global landscape of 80*S* ribosomes, we averaged the position-specific RPF densities of individual genes into a composite metagene, in which each gene was first normalized for its overall density of RPFs (i.e., RPKM of RPFs) and then weighted equally in the average (Equation S10). A small excess of ribosome density was observed in the first ~200 codons compared to the remainder of the ORF (Figure 1C). The trend toward decreasing ribosome density with codon position was also evident on a gene-by-gene basis: 82% of genes exhibited declining raw RPF reads along their entire gene-length, based on linear-regression of RPF reads with codon position (binomial test, $p < 10^{-15}$), with the 5′-to-3′ decrease in ribosome densities for a gene of average length (~500 codons) averaging ~43%.

Much larger 5′ ramps are observed in other studies (Ingolia et al., 2009; Gerashchenko et al., 2012; Zinshteyn and Gilbert, 2013; Artieri and Fraser, 2014; Guydosh and Green, 2014; McManus et al., 2014), which is attributed to their use of CHX pre-treatment (Ingolia et al., 2009; Gerashchenko and Gladyshev, 2014) (Figure S1). However, CHX pre-treatment cannot explain the more modest ramp observed in our dataset, since our protocol did not involve such treatment.

The 5′ ramp of ribosomes has previously been attributed to slower elongation due to preferential use of codons corresponding to low-abundance cognate tRNAs at the 5′ ends of genes (Tuller et al., 2010). To determine the contribution of codon usage, we tested whether differences in RPF densities between the 5′ and 3′ ends of genes depended on codon choice. Surprisingly, for each of the 61 sense codons, the average density of RPFs was 33% greater on average when the codon fell within

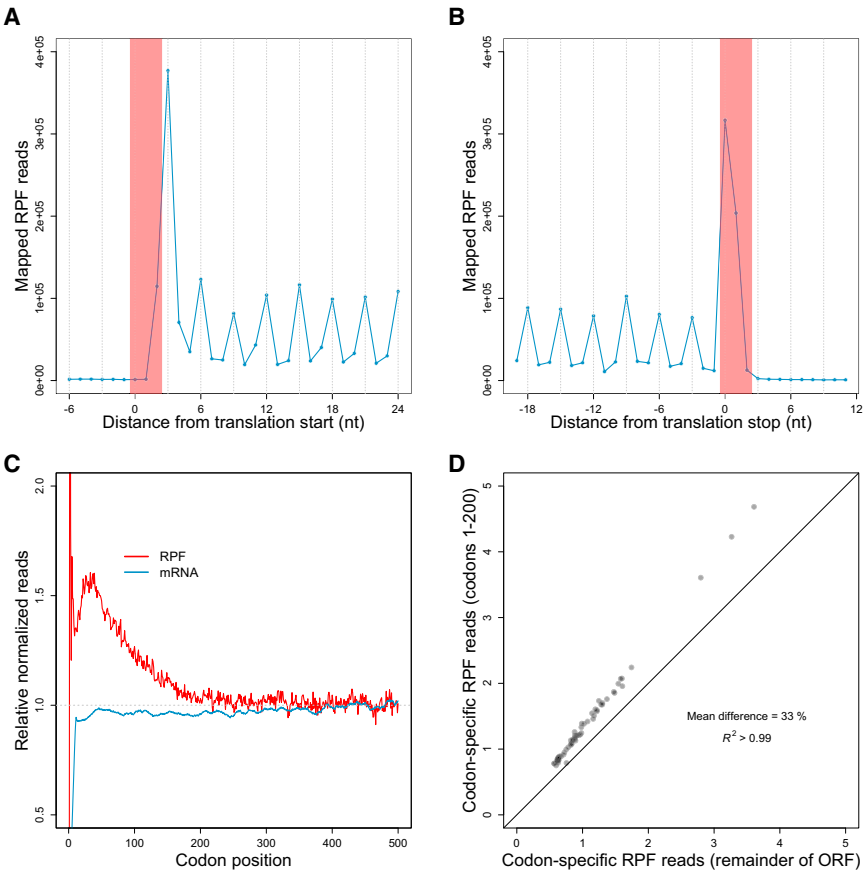


Figure 1. Less Perturbed RPFs Reveal a Codon-Independent 5′ Ramp

(A and B) Metagene analyses of RPFs. Coding sequences were aligned by their start (A) or stop (B) codons (red shading). Plotted are the numbers of 28–30-nt RPF reads with the inferred ribosomal A site mapping to the indicated position along the ORF.

(C) Metagene analyses of RPFs and RNA-seq reads (mRNA). ORFs with at least 128 total mapped reads between ribosome-profiling (red) and RNA-seq (blue) samples were individually normalized by the mean reads within the ORF, and then averaged with equal weight for each codon position across all ORFs (e'_j in Equation S10 and h'_j in Equation S14).

(D) Comparison of codon-specific RPFs as a function of the 5′ ramp. For each of the codons, densities of RPFs with ribosomal A sites mapping to that codon were calculated using either only the ramp region of each ORF (codons 1–200) or the remainder of each ORF (v^5_k in Equation S16 and v^3_k in Equation S17, respectively). The diagonal line indicates the result expected for no difference between the two regions.

See also Figure S1.

the first 200 codons of an ORF (Figures 1D and S1), which showed that differential codon usage alone cannot explain the 5′ ramp. Consistent with these experimental results, simulation of protein translation in a yeast cell, using a whole-cell stochastic model of yeast translation (Shah et al., 2013), indicated that codon ordering could account for at most a 20% ramp (Figure S1). Thus, codon ordering might explain some of the ~60% ramp observed in our dataset, but the majority of this ramp is likely caused by other mechanisms (see Discussion).

Codon-Specific Elongation Dwell Times Are Inversely Correlated with tRNA Abundances

The 61 sense codons varied in their average RPF densities by more than 6-fold (Figure 1D), indicating that different codons are decoded at different rates. Molecular biologists have long assumed that such differences in elongation rates are caused by corresponding differences in the cellular abundances of cognate tRNAs (Andersson and Kurland, 1990; Bulmer, 1991). Several early experiments provide empirical support for this view (Varenne et al., 1984; Sørensen and Pedersen, 1991), but early analyses of ribosome-profiling results do not find any relationship between ribosome density and cognate tRNA abundance expected from this model (Ingolia et al., 2011; Li et al., 2012; Qian et al., 2012; Charneski and Hurst, 2013; Zinshteyn and Gilbert, 2013). However, the datasets analyzed in these studies were all from experiments that used CHX pre-treatment.

At least three considerations help explain why CHX pre-treatment would disrupt the correlation between tRNA abundances and measured ribosome densities at the A site. The first is that CHX, once bound to a ribosome, allows for an additional round of elongation before halting ribosomes (Schneider-Poetsch et al., 2010; Gardin et al., 2014; Lareau et al., 2014), which alone would remove any correlation at the A site. Second, CHX binding is reversible, and at concentrations typically used in ribosome-profiling protocols, additional rounds of elongation might occur between CHX-binding events. Third, CHX prevents translocation of the ribosome by binding to the E site, with space for a deacylated tRNA (Schneider-Poetsch et al., 2010), and thus CHX binding affinity presumably varies with features of the E site and the tRNA in it. Thus, in the presence of CHX pre-treatment, the ribosome density at a site is likely more a function of the on and off rates of CHX binding than a function of differential isoaccepting tRNA availability. Indeed, recent analyses of profiling results obtained without CHX pre-treatment have observed modest correlations between tRNA abundances and ribosome-densities at the A site (Gardin et al., 2014; Lareau et al., 2014).

When examining earlier ribosome-profiling datasets, we found that whenever CHX pre-treatment was employed, the relationship between ribosome occupancy and tRNA abundance was both insignificant ($p > 0.05$) and in the opposite direction than expected (Figures S2C–S2E). Moreover, the concordance between these CHX pre-treatment datasets indicated a systematic bias (Figure S2), suggesting that an orthogonal set of mRNA sequence biases influence CHX binding. In contrast, for every dataset without CHX pre-treatment, we found that ribosome

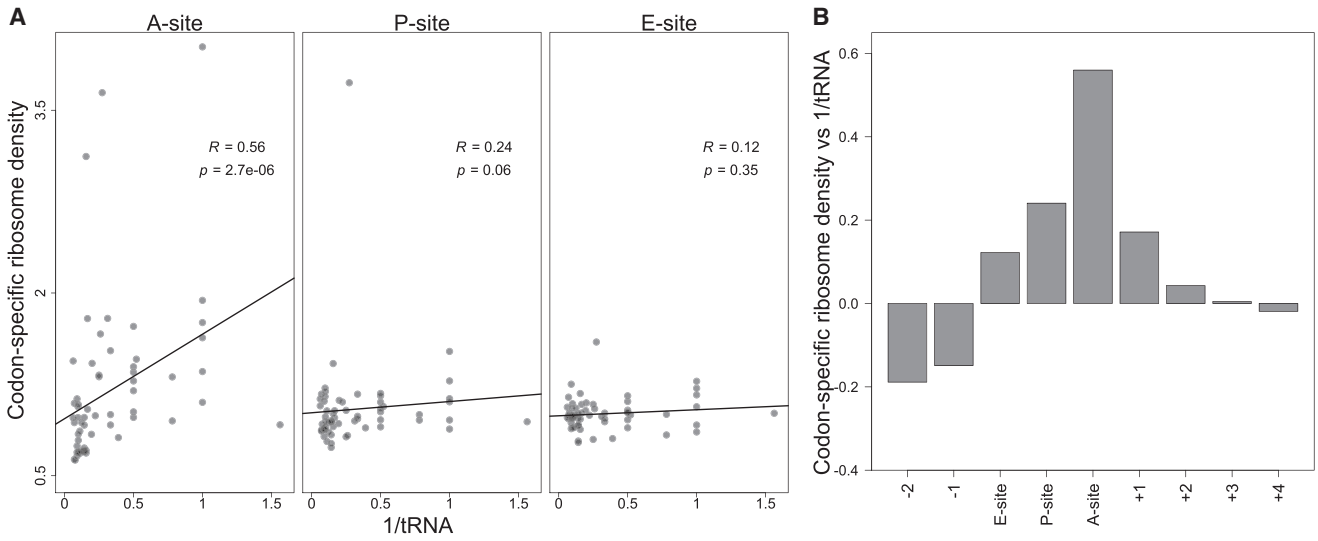


Figure 2. Codons Corresponding to Lower-Abundance tRNAs Are Decoded More Slowly

(A) Correlation between codon-specific excess ribosome densities and cognate tRNA abundances. Codons within RPFs were assigned to the A-, P-, and E-site positions based on the distance from the 5′ ends of fragments, and codon-specific excess ribosome densities were calculated (v_k in Equation S19). Cognate tRNA abundances for each codon were estimated using the genomic copy numbers of iso-accepting tRNAs and wobble parameters (Table S2). Spearman *R* values are shown, with their significance (p values).

(B) The correlations of codon–tRNA abundance at different positions relative to the A site. Analysis was as in (A) using varying offsets from the A-site position within RPFs (x axis) to calculate Spearman correlations (y axis).

See also Figures S2 and S3 and Tables S1 and S2.

densities were inversely correlated with tRNA abundances (Figures S2C–S2E).

In our dataset, we found that codon-specific excess ribosome densities (v_k in Equation S19) were strongly anti-correlated with cognate tRNA abundances, as estimated by copy numbers of tRNA genes and wobble parameters (Figures 2A and 2B). This strong anti-correlation was also observed with direct estimates of tRNA abundances obtained from our RNA-seq measurements (Figure S2A; Table S1). As expected, the correlation was specific to the codon within the A site, with residual correlations at the P and E sites, which were potentially caused by some 5′ heterogeneity of RPFs.

Taken together, these results strongly support the idea that differential cognate tRNA abundances influence differential elongation times of codons in the absence of CHX. Without CHX pretreatment, we also observed widespread pausing after polybasic tracts (Figure S3) but not at P-site proline codons (Figure S2), which has been the subject of some debate (Supplemental Information).

Slower Elongation at Regions Encoding Inter-domain Linkers

The modulation of elongation rates by either tRNA abundances (Figure 2A) or polybasic stretches (Figure S3) might influence the kinetics of co-translational folding. Indeed, slower elongation rates within inter-domain linkers relative to the adjacent domains is reported to coordinate co-translational folding of nascent polypeptides (Thanaraj and Argos, 1996; Kimchi-Sarfaty et al., 2007; Pechmann and Frydman, 2013). However, systematic experimental evidence for such differences in elongation rates has been lacking.

To examine whether our ribosome-profiling data reveal such differences, we first used InterProScan classifications (Jones et al., 2014) based on the Superfamily database (Wilson et al., 2009) to partition coding sequences into domain and linker regions. We then calculated the mean normalized RPF densities (z_{ij} in Equation S7) for codons within the domain- and linker-encoding regions and found significantly lower densities in regions of genes that fell within domains compared those that fell outside of domains (Figure 3; mean difference 0.094, paired t test, $p < 10^{-26}$). To eliminate any influence of the 5′ ramp, we repeated the analysis excluding the first 200 codons. Although the size of the effect diminished (mean diff = 0.029), the difference in mean ribosome densities remained significant (p = 0.0002), indicating that the 5′ ramp was not solely responsible for lower ribosome densities within domains (Figure S4A).

The trend toward relatively lower ribosome densities in domain regions held even when restricted to each individual amino acid, with the exceptions of cysteine residues and the single-codon-encoded methionine and tryptophan residues (Figure S4). Thus, differences in amino acid content between domains and linkers could not account for the observed differences in bound ribosome densities. Moreover, for 54 out of 61 sense codons, we found significantly lower ribosome densities in domains compared to linkers (one-sided t test, $p < 0.05$). For 26 out of 61 codons, we found significantly lower ribosome densities in domains even after excluding the first 200 codons (one-sides t test, $p < 0.05$). This result implied that differences in synonymous codon usage between domain and linker regions cannot alone

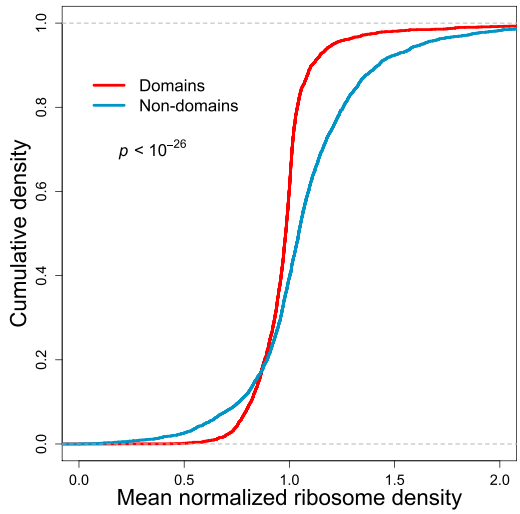


Figure 3. Elongation Dynamics Correlate Domain Architecture

Cumulative distributions of normalized ribosome densities within and outside of protein-folding domains. Mean normalized RPF densities (z_{ij} in Equation S7) for codons within the domain-encoding and non-domain-encoding regions were individually calculated for each ORF. Domain assignments were based on InterProScan classifications (Jones et al., 2014) obtained from the Superfamily database (Wilson et al., 2009). Statistical significance was evaluated using paired t test ($p < 10^{-26}$).

See also Figure S4.

account for the differences in ribosome densities. One possible mechanism for differential ribosome occupancy, independent of codon usage, is differential recruitment of chaperones and their associated effects on co-translational folding (Ingolia, 2014).

Similar results for densities in domain and linker regions were obtained when using InterProScan classifications (Bateman et al., 2002) instead of the Superfamily database (Figure S4B). Finally, consistent with other computational analyses (Pechmann and Frydman, 2013), differences in elongation rate were found at the level of protein secondary structures as well: regions corresponding to helices and sheets exhibited significantly lower RPF densities than regions corresponding to loops (Figure S4C). Taken together, these results provided systematic empirical support for the claim that co-translational folding requirements influence elongation rates. Nonetheless, the magnitude of this signal was very small, suggesting that slower inter-domain elongation either has very little impact or impacts very few genes.

Estimates of Protein-Synthesis Rates

Our results thus far indicated that the ribosome density at a given codon position is influenced by the abundance of cognate tRNAs and whether the codon is immediately downstream of a polybasic stretch, falls within a protein domain, or lies in the 5′ region of the ORF. The non-uniform ribosome densities along individual ORFs imply that the overall RPF density on each gene (i.e., RPKM of RPFs) does not directly reflect the rate of protein synthesis (Li et al., 2014). For example, the RPF densities of genes enriched in more slowly elongated codons would tend to overestimate their protein-synthesis rates, and the same would be true for shorter ORFs. To more accurately quantify the protein-synthesis rates of individual genes from RPF densities, we used empirically derived correction factors to account for the position- and codon-specific effects we observed (f_j in Equation S23). With these correction factors, the ~74.3 million sequenced RPFs enabled reliable estimates of protein-synthesis rates for 4,839 genes (Equation S28).

Accurate Measurement of Yeast mRNA Abundances

In addition to improving measurements of ribosome densities, we sought to improve measurements of mRNA abundances, which is also critical for accurately quantifying translational control. Prior experiments have typically measured yeast mRNA abundances by performing RNA-seq on poly(A)-selected RNA (Ingolia et al., 2009; Gerashchenko et al., 2012; Zinshteyn and Gilbert, 2013; Artieri and Fraser, 2014; Guydosh and Green, 2014; McManus et al., 2014). However, poly(A) selection might bias mRNA-abundance measurements. For example, mRNAs that lack a poly(A) tail of sufficient length to stably hybridize to oligo(dT) might not be as efficiently recovered. Although *S. cerevisiae* is not known to contain translated mRNAs that altogether lack a poly(A) tail, the lengths of poly(A) tails found on *S. cerevisiae* mRNAs are relatively short, with a median length of 27 nt (Subtelny et al., 2014). Another source of potential bias in poly(A) selection is partial recovery of mRNAs endonucleolytically cleaved during RNA isolation or poly(A) selection. The 5′ fragments resulting from mRNA cleavage are not recovered by poly(A) selection, which causes a 3′ bias in the resulting RNA-seq data (Nagalakshmi et al., 2008). Indeed, analyses of published RNA-seq datasets from ribosome-profiling studies revealed a severe 3′ bias in poly(A)-selected RNA-seq reads, ranging from 19%–130% excess reads (Equation S15) (Figure S5). Because longer mRNAs have a higher probability of being cleaved, the abundances of longer mRNAs might be systematically underestimated by poly(A) selection (Table S3).

An alternative to poly(A) selection is rRNA depletion, which enriches mRNAs by removing rRNA using subtractive hybridization. A concern with subtractive hybridization is the potential depletion of mRNAs that either cross-hybridize to the oligonucleotides used to remove rRNA sequences or adhere to the solid matrix to which the oligonucleotides are attached. To investigate the extent to which unintended mRNA depletion occurs when using reagents sold for yeast RNA-seq library preparations, we subjected the same total RNA to each of three procedures: Dynabeads oligo$(dT)_{25}$ (Life Technologies), RiboMinus Yeast Transcriptome Isolation Kit (Life Technologies), or Ribo-Zero Yeast Magnetic Gold Kit (Epicenter). As a reference, we also generated an RNA-seq library from the total RNA that was not enriched for mRNA and thus contained primarily rRNA (90.2% of 199.7 million genome-mapping reads). We also note that we started with RNA extracted from the lysate that was used for ribosome-footprint profiling, as opposed to RNA extracted from whole cells as done in the original ribosome-profiling study (Ingolia et al., 2009). When comparing the 4,540 mRNAs for which we obtained at least 64 reads in our total RNA library, only the Ribo-Zero-treated sample faithfully recapitulated the mRNA abundances observed in total RNA (R^2 = 0.98; Figures 4A and S5). The poly(A)-selected and RiboMinus-treated

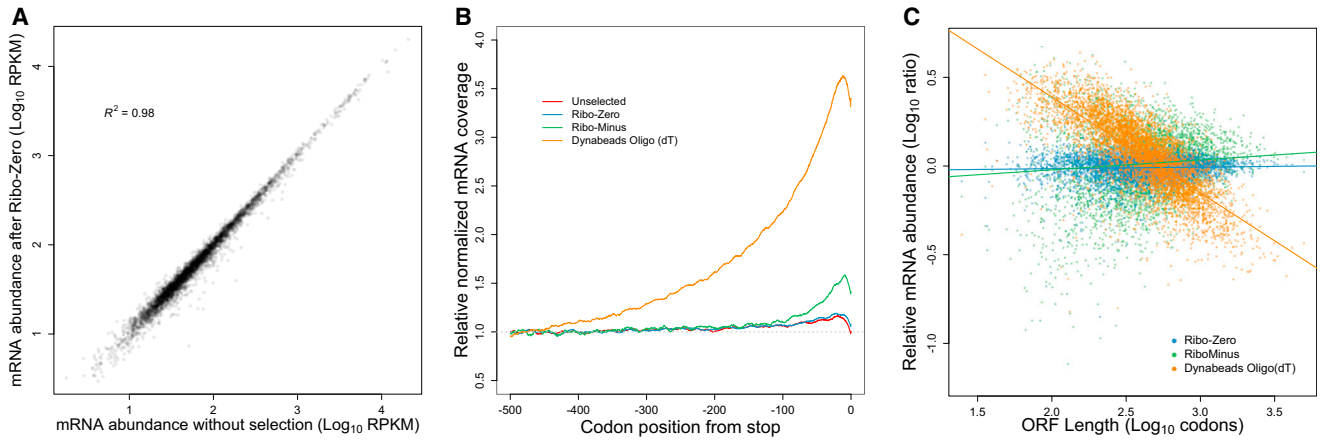


Figure 4. mRNA Enrichment Methods Can Bias mRNA Abundance Measurements

(A) mRNA abundances measured by RNA-seq of Ribo-Zero-treated RNA compared to those measured by RNA-seq of total unselected RNA. Pearson R^2 is indicated.

(B) Metagene analysis of RNA-seq read density in total unselected or mRNA-enriched RNA samples. Coding sequences were aligned by their stop codons, and RNA-seq reads were individually normalized by the mean reads within the ORF and then averaged with equal weight for each codon position across all ORFs (h''_j in Equation S15).

(C) mRNA abundances for mRNA-enriched samples relative to total unselected RNA, as a function of ORF length.

See also Figures S5, S6, and S7 and Tables S3 and S4.

samples each had significantly lower correlations with total RNA (R^2 = 0.85 and R^2 = 0.87, respectively), indicating a skewed representation of the transcriptome. Compared to RNA-seq data from published ribosome-profiling studies, our Ribo-Zero-treated sample also exhibited the highest correlations with microarray-based estimates of mRNA abundances (Table S3).

As anticipated, the poly(A)-selected sample contained a strong 3′ bias (Figure 4B), which caused a systematic underestimation of the abundances of longer genes (Figure 4C). After accounting for this strong bias in the poly(A)-selected sample, we did not detect a relationship between poly(A)-tail length and poly(A)-selection efficiency, suggesting that tail-length differences did not significantly contribute to the biases of poly(A)-selected RNA-seq data. For the RiboMinus-treated sample, cross-hybridization to the depletion probes might have skewed the mRNA abundances, which might have been largely avoided in the Ribo-Zero protocol because of its more stringent hybridization conditions. The RiboMinus-treated sample also had substantial rRNA contamination (44.5% of reads, originating primarily from the 5*S* rRNA).

Interestingly, the total-RNA and the Ribo-Zero datasets both contained a small 3′ bias (Figure 4B), with median 3′/5′ excess reads of 22% and 28%, respectively (Table S4). This bias was consistent with reports that yeast mRNAs are primarily degraded in the 5′-to-3′ direction (Hu et al., 2009; Pelechano et al., 2015). The decay intermediates of this vectorial degradation process would contribute more reads toward the 3′ ends of mRNAs, giving rise to the observed bias, especially when considering that our RNA samples were enriched for cytoplasmic RNA, which would diminish the countervailing vectorial mRNA synthesis process occurring in the nucleus. Nonetheless, the 3′ biases in the total-RNA and Ribo-Zero datasets were smaller than those in poly(A)-selected samples, for which median 3′/5′ excess mRNA reads ranged from 42% to 275% (Table S4). Because Ribo-Zero treatment enabled deep coverage of the yeast transcriptome without substantially biasing mRNA abundances, we used mRNA abundances estimated from Ribo-Zero-treated RNA for all subsequent analyses.

A Narrow Range of Initiation Efficiencies in Log-Phase Yeast

Because protein synthesis is typically limited by the rate of translation initiation (Andersson and Kurland, 1990; Bulmer, 1991; Shah et al., 2013), we defined the initiation efficiency (IE) of each gene as its protein-synthesis rate divided by its mRNA abundance (Equation S27). Thus, the IE measure quantified the efficiency of protein production per mRNA molecule of a gene, in a typical cell. To facilitate comparisons with published datasets, we also calculated the translational efficiency (TE) of each gene, defined as its RPF density normalized by its mRNA abundance (Ingolia et al., 2009). Because TE is calculated based on the RPF density rather than the protein-synthesis rate, TE does not account for differential rates of elongation associated with the 5′ ramp or codon identity. Nonetheless, IE and TE were highly correlated (R = 0.951; Figure S6A).

A wide range of IEs (or TEs) among genes would indicate that protein production is under strong translational control, whereas a narrow range would indicate that protein production is typically governed by mRNA abundances, and hence protein-synthesis rate is primarily controlled by mRNA transcription and decay. The first ribosome-profiling study suggested a large amount of translational control in yeast, with the range of TEs reported to span roughly 100-fold (Ingolia et al., 2009). Indeed, we found that the 1–99 percentile range of TEs in those data spanned 73-fold (Figure S6C). In contrast, the range of TEs observed in our data was narrower, with the 1–99 percentile spanning only

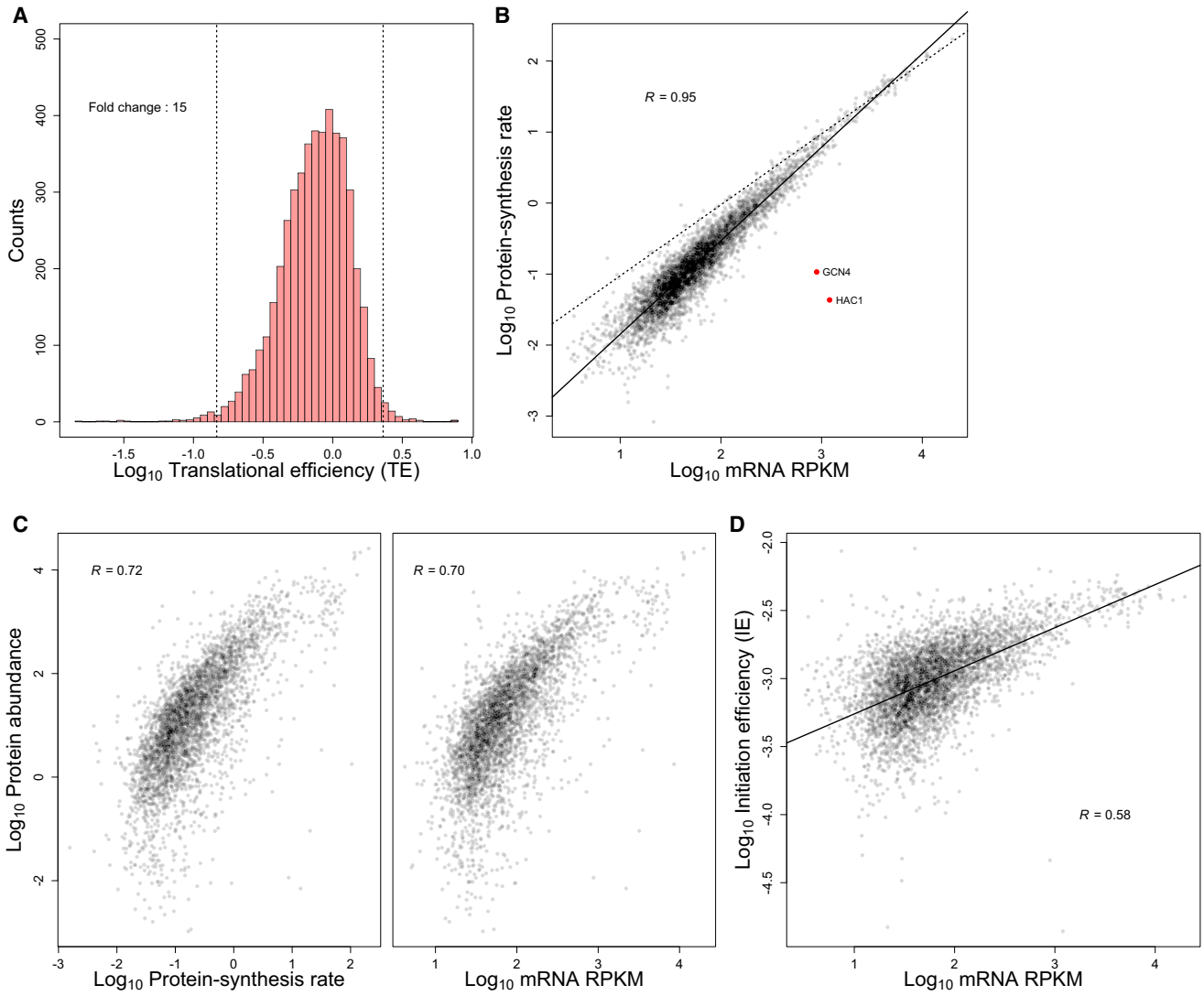


Figure 5. TEs and IEs Span a Narrow Range in Log-Phase Yeast Cells

(A) Distribution of TE measurements, with vertical dashed lines marking the first and 99th percentiles, and the fold change separating these percentiles indicated. All ORFs with at least 128 total reads between the ribosome-profiling and RNA-seq datasets were included (except *YCR024C-B*, which was excluded because it is likely the 3′ UTR of *PMP1* rather than an independently transcribed gene).

(B) Relationship between estimated protein-synthesis rate and mRNA abundance for genes shown in (A). *GCN4* and *HAC1* (red points) were the only abundant mRNAs with exceptionally low protein-synthesis rates. The best linear least-squares fit to the data is shown (solid line), with the Pearson *R*. For reference, a one-to-one relationship between protein-synthesis rate and mRNA abundance is also shown (dashed line).

(C) Relationship between experimentally measured protein abundance (de Godoy et al., 2008) and either protein-synthesis rate (left) or mRNA abundance (right). The 3,845 genes from (A) for which protein-abundance measurements were available were included in these analyses. Pearson correlations are shown (*R*).

(D) Relationship between mRNA abundance and IE for genes shown in (A). The best linear least-squares fit to the data is shown, with the Pearson *R*.

See also Figures S8 and S9 and Table S5.

a 15-fold range (Figure 5A). Although the range of IEs was marginally wider than that of TEs (1–99 percentile spanning 21-fold; Figure S6B), it was still substantially smaller than the range of TEs initially reported (Ingolia et al., 2009). The relatively narrow range of IEs in our data was also reflected by the high correlation between mRNA abundance and protein-synthesis rate (R = 0.948; Figure 5B), supporting the conclusion that protein-synthesis rates are largely dictated by mRNA abundances (Csárdi et al., 2015). Interestingly, the slope of the regression between mRNA and protein-synthesis rates was >1 on the log-scale, indicating that translation regulation mostly amplifies the effect of differential mRNA abundances rather than buffering it (Csárdi et al., 2015). Further indicating that mRNA abundance (when accurately measured) is a strong predictor of total protein production, mass-spectrometry-based measurements of steady-state protein abundance (de Godoy et al., 2008) correlated as well with mRNA abundances as they did with protein-synthesis rates (Figure 5C).

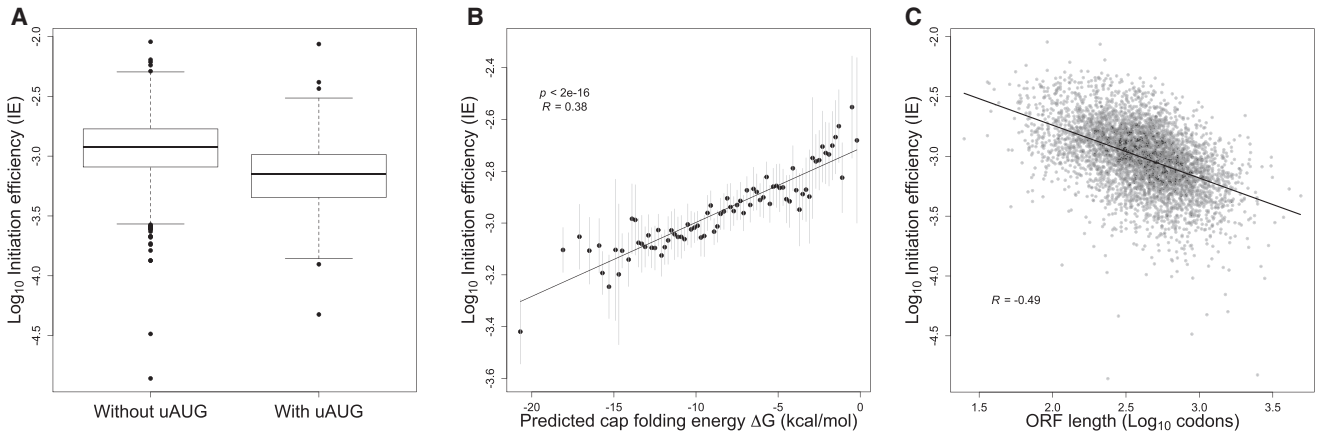


Figure 6. mRNA Sequence, Structure, and Length Correlate with IE

(A) Reduced IE values for genes with at least one upstream AUG (i.e., an AUG codon located within the annotated 5′ UTR). The plots indicated the median (line), quartile (box) and first and 99th percentiles (whiskers) of the distributions.

(B) Inverse relationship between IE and the folding energy of predicted RNA secondary structure near the cap (Cap-folding energy). RNAfold was used to estimate folding energies for the first 70 nt of the mRNA. Gray bars indicate 1 SD of IE values for genes binned by predicted folding energy. The best linear least-squares fit to the data is shown (solid line), with the Pearson *R*.

(C) Inverse relationship between IE and ORF length. The best linear least-squares fit to the data is shown (solid line), with the Pearson *R*.

See also Figure S7.

When we examined the range of TEs in other published datasets, we also found more narrow ranges (as low as 11-fold from 1–99 percentiles) than that of Ingolia et al. (2009) (Figure S6C). However, the TEs in published datasets—which are all generated using poly(A)-selected mRNA—were not particularly well correlated with each other (Table S5). These discrepancies in TEs were largely due to differences in measured mRNA abundances, whereas the RPF abundances correlated almost perfectly (Table S5). Collectively, these results indicate that the amount of translational control in log-phase yeast has been overestimated due to inaccuracies in TE measurements, largely caused by challenges in accurately measuring mRNA levels.

We also noticed that the shape of the TE distribution from our data, which was asymmetric, differed from that of the Ingolia data, which is highly symmetric. In particular, in our data there were relatively few genes in the right tail of the distribution (Figure 5A; note the location of the mode closer to the 99th than the first percentile). This observation implied that mRNAs from very few genes contain elements that impart an exceptionally high initiation efficiency and are thereby "translationally privileged." Rather, most mRNAs either initiate close to a maximum possible rate (likely set by the availability of free ribosomes or initiation factors) or contain features that modestly reduce the initiation rate.

To the extent that differences in IE were observed, the genes with lower IE tended to be expressed at lower mRNA levels, with IE increasing roughly linearly with mRNA expression levels (Figure 5D). These results were consistent with the notion that abundant mRNAs have undergone evolutionary selection to be efficiently translated (Sharp and Li, 1987; Andersson and Kurland, 1990; Plotkin and Kudla, 2011; Shah and Gilchrist, 2011). Interestingly, in the plots comparing protein-synthesis rate or IE with mRNA level, the points for 11 of the 12 highest expressed mRNAs fell below the regression lines (Figures 5B and 5D, dashed lines), suggesting that the efficiency for the highest expressed mRNAs might have saturated.

Two notable outliers appeared in the comparison of mRNA abundances and synthesis rates (Figure 5B, red dots). These two, which corresponded to relatively abundant mRNAs with exceptionally low synthesis rates, were *HAC1* and *GCN4*. These are the two most well-known examples of translational control in log-phase yeast and are both involved in rapid stress responses, during which translational repression is relieved (Rüegsegger et al., 2001; Mueller and Hinnebusch, 1986; Dever et al., 1992). The observation that *HAC1* and *GCN4* were the only abundant mRNAs that were strongly regulated at the translational level further emphasized that translational control only modestly influences the protein production of most yeast genes. Nevertheless, the tuning of synthesis rates via translational control can help maintain the proportional synthesis of the subunits of multiprotein complexes (Figures S6D–S6G; Supplemental Experimental Procedures).

Determinants of Initiation Efficiencies in Yeast

Next, we sought to identify sequence-based features that explain the variation in IE values that remained among genes after improving the RPF and mRNA measurements. First, we considered uORFS, which can inhibit translation by serving as decoys to prevent initiation at the start codons of bona fide ORFs (Zur and Tuller, 2013), as occurs for *GCN4* (Mueller and Hinnebusch, 1986; Dever et al., 1992), one of two genes with the greatest translational repression (Figure 5B). Using high-resolution 5′ UTR annotations (Arribere and Gilbert, 2013), we identified upstream AUGs (uAUGs) in 303 out of the 2,549 genes that had reproducibly uniform transcription-start sites. Those genes containing uAUGs had significantly lower IEs than genes without uAUGs, even after controlling for 5′ UTR lengths (Figure 6A; t test $p < 10^{-16}$). These results confirmed that a general feature of

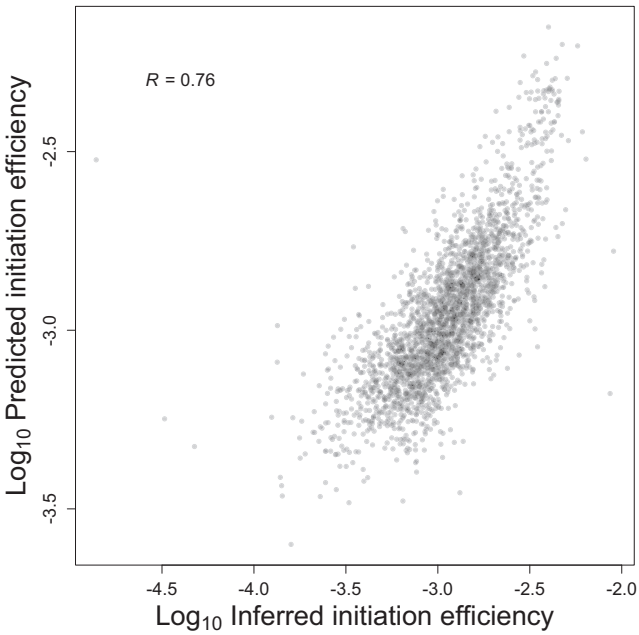


Figure 7. Sequence-Based Features of mRNAs Largely Explain Yeast IEs

Correspondence between predicted IEs and IEs inferred directly from the RPF and RNA-seq data. Initiation efficiencies were predicted using a multiple-regression model, based on mRNA abundance and sequence-based features of the 2,549 genes with empirically determined 5′ UTRs. Shown is the Pearson *R*. See also Table S6.

uORFs is to decrease the translation of downstream ORFs, and that the presence of uAUGs can explain some of the variance in IEs (Arribere and Gilbert, 2013; Zur and Tuller, 2013).

Another feature that has been linked to differences in synthesis rates is mRNA secondary structure. Structure located near the 5′ cap might interfere with binding of the eIF4F cap-binding complex, while structure within the 5′ UTR could disrupt the scanning 40*S* ribosome. An open structure around the start codon might also be important for facilitating joining of the 60*S* subunit. Previous genome-wide structure analyses revealed a weak but significant inverse correlation between start-codon-proximal structure and TE (Kertesz et al., 2010), but the accessibility of the 5′ UTR more generally was not reported, and the TE values used in those studies were affected by RNA-seq biases. For each mRNA with a single reproducible 5′ end (Arribere and Gilbert, 2013), we predicted the accessibility of the 5′ cap by calculating the predicted folding energy of the sequence spanning increasing distances from the cap. For all distances examined, we observed a significant correlation between predicted cap accessibility and IE (t test, $p < 10^{-6}$ for each window; Figures 6B and S7). This correlation rapidly increased with window length, approaching a maximum at 70–90 nt (Pearson correlation, $R \sim 0.37$ for windows 70–90 nt) and then steadily declined for larger windows (Figure S7), consistent with local folding of the 5′ end determining cap accessibility. Together, these results confirmed that mRNAs with less-structured 5′ UTRs tend to be initiated more efficiently (Godefroy-Colburn et al., 1985; Shah et al., 2013), which is consistent with eIF4F binding, 40*S* recruitment, or scanning as influential regulatory steps during eukaryotic initiation. Notably, the correlations that we observed between predicted mRNA structure and translation were the largest that have been reported between these features in eukaryotes, which emphasized the utility of our accurate IE measurements and of predicting structure near the cap as opposed to more downstream regions.

Gene length has also been reported to correlate with translational efficiency. Although global polysome-profiling studies indicate strong anti-correlation between ORF length and ribosome density (Arava et al., 2003), analysis of published ribosome-footprint-profiling data revealed essentially no correlation (or even a positive correlation in some cases) between length and TE (Figure S7). In contrast, we observed a striking negative correlation in our IE (and correspondingly in our TE) data (Figures 6C and S7). Our IE measure already corrected for the elevated ribosome densities in the first 200 codons, and the negative correlation between ORF length and TE persisted even after removing the first 250 codons of each ORF, which further confirmed that this correlation was not caused by the 5′ ramp (Figure S7). The discrepancy between our data and earlier ribosome-profiling datasets was likely due to the RNA-seq 3′-bias caused by poly(A) selection (Figures 4B and S5). Indeed, an anti-correlation between ORF length and TE was observed in most other datasets when we controlled for the 3′ bias by estimating mRNA abundances based on mapped RNA-seq reads from only the 3′ ends of genes (Figure S7). Together, these results showed that the original report of shorter mRNAs having relatively higher initiation efficiencies (Arava et al., 2003) is correct, even after accounting for the CHX-enhanced 5′ ramp that confounded that analysis.

A Statistical Model that Predicts Initiation Efficiencies

Based on these results, we used multiple linear regression to build a model that considered number of uAUGs, predicted cap-proximal RNA-folding energy (and also GC content of the 5′ UTR as another metric for structure), and lengths of the ORF and the 5′ UTR to explain the variance in IE observed among genes. We also included an mRNA-abundance term in the model because IE is greater for more abundant mRNAs (Figure 5D). To identify the most informative features, we used Akaike's Information Criteria (AIC) for model selection and both step-up and step-down model-selection procedures (using the stepAIC function in the MASS package in R). The multiple regression model that best explained the variation in IE included all six variables, even after penalizing for model complexity (Figure 7; Table S6). The dominant explanatory variable was mRNA abundance, which alone accounted for ~40% of the variance in IE. Collectively, a model containing all six variables explained ~58% of the variance in IE. A model that excluded mRNA abundance, and therefore depended on only sequence-based features, still explained ~39% of the variance in IE. These results of our statistical modeling should help motivate mechanistic studies of how each of these mRNA features impacts translation.

DISCUSSION

We have shown that improved measurements of both mRNA abundances and RPFs can provide insights into the regulation

and dynamics of eukaryotic translation. The RPFs that we isolated and sequenced are indicative of a dynamic and heterogeneous elongation process, with ribosomes transiting along mRNA molecules at variable rates depending on the distance from the start codon, codon identity, and nascent polypeptide sequence.

What might explain the 5′ ramp of ribosomes observed even in the absence of CHX pre-treatment (Figure 1C)? Codon usage accounted for about a third of it, but even the same codons were differentially occupied by ribosomes depending upon whether they occurred in the 5′ or 3′ ends of genes (Figure 1D), indicating that additional mechanisms must be involved. Although we cannot rule out ribosome drop-off as a contributing factor, we favor the idea that elongation is slower during the early phase of translation. Perhaps an initiation factor remains engaged with the 80*S* ribosome during early elongation, and the bound factor maintains the ribosome in a slower state until it stochastically dissociates from the ribosome within the first 200 codons. The eIF3 complex is a promising candidate for such a factor, as it binds the solvent-exposed face of the 40*S* ribosome (Siridechadilok et al., 2005) and can therefore bind to 80*S* ribosomes as well (Beznosková et al., 2013). Maintaining eIF3 on early elongating ribosomes might also facilitate re-initiation after translation of short uORFs (Szamecz et al., 2008; Zur and Tuller, 2013).

A practical finding of our studies is that the choice of mRNA enrichment method can have a significant impact on yeast mRNA-abundance measurements. rRNA depletion using the Ribo-Zero kit was the only method that enriched for mRNAs without introducing substantial and systematic biases (Figures 4A and S5). One caveat of rRNA depletion is that nascent pre-mRNAs that lack a poly(A) tail may also be recovered, which can inflate mRNA abundance measurements with respect to the pool of translatable mRNA molecules. This effect may be more pronounced in metazoans that contain long introns and correspondingly long transcription times. The extent to which poly(A)-selection biases affect metazoan mRNA abundance data and thereby influence TE measurements remains to be determined.

The initial report that TE spans a roughly 100-fold range across mRNAs in budding yeast spurred intensive investigation of the underlying TE determinants, with varying degree of success (Kertesz et al., 2010; Tuller et al., 2011; Charneski and Hurst, 2013; Zur and Tuller, 2013; Bentele et al., 2013; Rouskin et al., 2014). Our results showed that this apparently wide range of TEs is partly explained by inaccurate mRNA-abundance measurements. After identifying and minimizing this source of inaccuracy, we observed a narrower range of TEs and IEs (Figure 5A; Table S3), suggesting a more limited degree of translational control. The TE range that we observed in yeast resembled the range observed in mouse embryonic stem cells (Ingolia et al., 2011), suggesting that limited translational control is a general principle of gene regulation in rapidly dividing eukaryotic cells.

Using our IE measurements, we were able to generate a statistical model that explained a majority of the IE variance (Figure 7; Table S6). Based on this model, secondary structure within the 5′ UTR, most especially cap-proximal structure, appears to be an important determinant of IE. These results are in agreement with early mechanistic studies demonstrating that cap accessibility correlates with initiation efficiency (Godefroy-Colburn et al., 1985) and that stable 5′ UTR secondary structures block the scanning ribosome (Kozak, 1986). One caveat of our structure analyses is that we used in silico prediction of mRNA structure, which does not always accurately capture the in vivo structure of mRNA (Rouskin et al., 2014). Further indicating the inadequacy of in silico predictions was the benefit of also including 5′ UTR GC content as a feature in our model. Likewise, the inclusion of mRNA abundance might have helped compensate for the inadequacy of in silico structure predictions, as highly expressed genes have less predicted structure in 5′ UTRs than do lowly expressed genes (Gu et al., 2010), and presumably these differences would be even greater when looking at actual 5′ UTR structure. Therefore, mRNA structure presumably explains even more variation in IE than our analyses suggested.

We also found that longer ORFs tended to be more poorly translated in log-phase yeast, even after accounting for the 5′ ramp (Figure 6C). Given that initiation occurs at the 5′ ends of mRNAs, how might initiation rates be sensitive to ORF lengths? One possibility is that shorter mRNAs, which include ribosomal proteins and other housekeeping genes (Hurowitz and Brown, 2003), might be under selection for faster initiation rates by virtue of their high expression. However, our stepwise regression showed that ORF length was informative even after accounting for mRNA abundance. Another possibility is that the 5′-UTR-bound initiation machinery can sense and be affected by ORF length via the closed-loop structure. In eukaryotes, translating mRNAs are thought to adopt a pseudo-circularized structure in which the 5′ and 3′ ends are in close proximity, enhancing translation and mRNA stability (Christensen et al., 1987). Previous biochemical analysis of the closed loop in yeast extracts revealed that only short mRNAs adopt a stable closed-loop structure in vitro (Amrani et al., 2008), presumably due to the relatively short distance between the mRNA termini. If the same principle applies in vivo, then inefficient closed-loop formation of long mRNAs could explain their relatively low IEs.

EXPERIMENTAL PROCEDURES

Yeast Culture, Harvesting, and Lysate Preparation

S. cerevisiae strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was grown at 30°C in 500 ml YPD to OD_{600} 0.5. Cells were harvested by filtration using a Kontes Ultra-Ware Microfiltration Assembly with a Supor 450 Membrane Disc Filter that had been pre-wet with YPD. As the last liquid flowed through, the filtration apparatus was rapidly disassembled, cells were gently scraped off of the filter using a cell lifter, and the scraper was immediately submerged in a 50-ml conical tube filled with liquid nitrogen. Once the liquid nitrogen had boiled off, the pellet was stored in the conical tube at –80°C until lysis. To lyse cells under cryogenic conditions, the cell pellet was transferred into a pre-chilled mortar that was surrounded and filled with liquid nitrogen. The pellet was ground to a fine powder with a pre-chilled pestle, transferred into a 50-ml conical tube filled with liquid nitrogen, and after boiling off the liquid stored at –80°C. Crude lysate was prepared by briefly thawing the cell powder on ice for 1 min and then resuspending in 4 ml polysome lysis buffer (10 mM Tris-HCl [pH 7.4], 5 mM $MgCl_2$, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 μg/ml cycloheximide, 500 U/ml RNasin Plus RNase Inhibitor [Promega], cOmplete EDTA-free Protease Inhibitor Cocktail [Roche]). The lysate was centrifuged at 1,300 × *g* for 10 min, and the supernatant was flash frozen in single-use aliquots.

RNA-Seq

Total RNA was extracted from an aliquot of frozen yeast lysate using TRI Reagent (Ambion) according to the manufacturer's protocol. Aliquots of the same sample were subjected to either no enrichment (the total RNA sample), poly(A) selection using 30 μg total RNA and 100 μl Dynabeads oligo(dT)$_{25}$ (Life Technologies) according to the manufacturer's instructions, rRNA depletion using 4 μg total RNA and the RiboMinus Yeast Transcriptome Isolation Kit (Life Technologies) according to the manufacturer's instructions, and rRNA depletion using 10 μg total RNA and the Ribo-Zero Gold Yeast rRNA Removal Kit (Illumina) according to the manufacturer's instructions. RNA samples were then diluted to 90 μl with water and precipitated with 10 μl 3 M NaCl, 30 μg GlycoBlue (Life Technologies), and 250 μl ethanol. RNA-seq was performed as described (Subtelny et al., 2014), using five cycles of PCR.

Ribosome Profiling

RPFs were isolated from an aliquot of frozen yeast lysate and sequenced on the Illumina HiSeq platform, as described (Subtelny et al., 2014). Detailed protocols for RNA-seq and ribosome profiling are available at http://bartellab.wi.mit.edu/protocols.html. RNase I treatment was performed using 0.2 U/μl lysate. Subtractive hybridization to remove contaminating rRNA fragments was performed using a mixture of three biotinylated oligonucleotides (Integrated DNA Technologies): 5′-GATCGGTCGATTGTGCACCTC/3Bio/; 5′-CGCTTCATTGAATAAGTAAAG/3Bio/; 5′-GACGCCTTATTCGTATCCATC/3Bio/.

Analyses

Equations and detailed procedures for analyses are provided in Supplemental Experimental Procedures.

ACCESSION NUMBERS

Sequencing data have been deposited in the GEO database under accession number GEO: GSE75897.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.01.043.

AUTHOR CONTRIBUTIONS

D.E.W., P.S., S.W.E., J.B.P., and D.P.B. designed the study. D.E.W., P.S., J.B.P., and D.P.B. wrote the manuscript, with help from the other authors. D.E.W. and S.W.E. prepared RNA-seq and ribosome-profiling libraries, respectively, under the supervision of D.P.B. D.E.W. performed initial data analysis. P.S. performed detailed data analyses and simulations with help from J.B.P. J.A.H. contributed new analytical tools.

ACKNOWLEDGMENTS

We thank J. Weissman and L. Lareau for helpful discussions, and the Whitehead Genome Technology Core for sequencing. This work was supported by the UCSF Program for Breakthrough Biomedical Research (funded in part by the Sandler Foundation, D.E.W.) and by NIH grants DP5OD017895 (D.E.W.) and GM061835 (D.P.B.), the Burroughs Wellcome Fund (J.B.P.), the David and Lucile Packard Foundation (J.B.P.), US Department of the Interior Grant D12AP00025 (J.B.P.), and US Army Research Office Grant W911NF-12-1-0552 (J.B.P.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

Received: July 6, 2015
Revised: November 17, 2015
Accepted: January 8, 2016
Published: February 11, 2016

REFERENCES

Amrani, N., Ghosh, S., Mangus, D.A., and Jacobson, A. (2008). Translation factors promote the formation of two states of the closed-loop mRNP. Nature *453*, 1276–1280.

Andersson, S.G., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. Microbiol. Rev. *54*, 198–210.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *100*, 3889–3894.

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res. *23*, 977–987.

Artieri, C.G., and Fraser, H.B. (2014). Evolution at two levels of gene expression in yeast. Genome Res. *24*, 411–421.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic Acids Res. *30*, 276–280.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. Mol. Syst. Biol. *9*, 675.

Beznosková, P., Cuchalová, L., Wagner, S., Shoemaker, C.J., Gunišová, S., von der Haar, T., and Valášek, L.S. (2013). Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. PLoS Genet. *9*, e1003962.

Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C.C., Li, G.W., Zhou, S., King, D., Shen, P.S., Weibezahn, J., et al. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell *151*, 1042–1054.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science *335*, 552–557.

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. Genetics *129*, 897–907.

Charneski, C.A., and Hurst, L.D. (2013). Positively charged residues are the major determinants of ribosomal velocity. PLoS Biol. *11*, e1001508.

Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am. J. Anat. *178*, 1–10.

Chu, D., and von der Haar, T. (2012). The architecture of eukaryotic translation. Nucleic Acids Res. *40*, 10098–10106.

Csárdi, G., Franks, A., Choi, D.S., Airoldi, E.M., and Drummond, D.A. (2015). Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. PLoS Genet. *11*, e1005206.

de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Fröhlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature *455*, 1251–1254.

Dever, T.E., Feng, L., Wek, R.C., Cigan, A.M., Donahue, T.F., and Hinnebusch, A.G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. Cell *68*, 585–596.

Ding, Y., Shah, P., and Plotkin, J.B. (2012). Weak 5′-mRNA secondary structures in short eukaryotic genes. Genome Biol. Evol. *4*, 1046–1053.

Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. eLife *2*, e01179.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. eLife *3*, 3.

Gerashchenko, M.V., and Gladyshev, V.N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res. *42*, e134.

Gerashchenko, M.V., Lobanov, A.V., and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc. Natl. Acad. Sci. USA *109*, 17394–17399.

Godefroy-Colburn, T., Ravelonandro, M., and Pinck, L. (1985). Cap accessibility correlates with the initiation efficiency of alfalfa mosaic virus RNAs. Eur. J. Biochem. *147*, 549–552.

Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. *4*, 117.

Gu, W., Zhou, T., and Wilke, C.O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput. Biol. *6*, e1000664.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guydosh, N.R., and Green, R. (2014). Dom34 rescues ribosomes in 3′ untranslated regions. Cell *156*, 950–962.

Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Coller, J. (2009). Co-translational mRNA decay in Saccharomyces cerevisiae. Nature *461*, 225–229.

Hurowitz, E.H., and Brown, P.O. (2003). Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. Genome Biol. *5*, R2.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science *346*, 1257521.

Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res. *39*, e141.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. Nature *467*, 103–107.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525–528.

Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA *83*, 2850–2854.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science *324*, 255–258.

Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. eLife *3*, e01257.

Li, G.W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. *24*, 422–430.

Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. Cell *45*, 201–207.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

Nedialkova, D.D., and Leidel, S.A. (2015). Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. Cell *161*, 1606–1618.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat. Struct. Mol. Biol. *20*, 237–243.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400–1412.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42.

Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol. Syst. Biol. *10*, 770.

Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. *8*, e1002603.

Richter, J.D., and Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature *433*, 477–480.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Rüegsegger, U., Leber, J.H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell *107*, 103–114.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat. Chem. Biol. *6*, 209–217.

Shah, P., and Gilchrist, M.A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. Proc. Natl. Acad. Sci. USA *108*, 10231–10236.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. Cell *153*, 1589–1601.

Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. *15*, 1281–1295.

Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., and Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science *310*, 1513–1515.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731–745.

Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V., and Dalmay, T. (2012). Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence *3*, 4.

Sørensen, M.A., and Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. J. Mol. Biol. *222*, 265–280.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Szamecz, B., Rutkai, E., Cuchalová, L., Munzarová, V., Herrmannová, A., Nielsen, K.H., Burela, L., Hinnebusch, A.G., and Valásek, L. (2008). eIF3a cooperates with sequences 5′ of uORF1 to promote resumption of scanning by

post-termination ribosomes for reinitiation on GCN4 mRNA. Genes Dev. *22*, 2414–2425.

Thanaraj, T.A., and Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. Protein Sci. *5*, 1594–1612.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Ruppin, E., and Ziv-Ukelson, M. (2011). Composite effects of gene determinants on the translation speed and density of ribosomes. Genome Biol. *12*, R110.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. *180*, 549–576.

Williams, C.C., Jan, C.H., and Weissman, J.S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science *346*, 748–751.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY–sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. *37*, D380–D386.

Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., and Johnson, C.H. (2013). Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature *495*, 116–120.

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature *495*, 111–115.

Zinshteyn, B., and Gilbert, W.V. (2013). Loss of a conserved tRNA anticodon modification perturbs cellular signaling. PLoS Genet. *9*, e1003675.

Zur, H., and Tuller, T. (2013). New universal rules of eukaryotic translation initiation fidelity. PLoS Comput. Biol. *9*, e1003136.